Zhang Y, Huang Y, Li N, Chambers JA.

A Robust Gaussian Approximate Fixed-Interval Smoother for Nonlinear Systems with Heavy-Tailed Process and Measurement Noises.

IEEE Signal Processing Letters 2016, 23(4), 468-472

A robust Gaussian approximate fixed-interval smoother for nonlinear systems with heavy-tailed process and measurement noises

Yulong Huang, Yonggang Zhang, *Member, IEEE,* Ning Li, Jonathon Chambers, *Fellow, IEEE*

***Abstract*—In this letter, a robust Gaussian approximate fixed-interval smoother for nonlinear systems with heavy-tailed process and measurement noises is proposed. The process and measurement noises are modelled as stationary Student's t distributions and the state trajectory and noise parameters are inferred approximately based on the variational Bayesian approach. Simulation results show the efficiency and superiority of the proposed smoother as compared with existing smoothers.**

***Index Terms*—Gaussian approximate smoother, Student's t distribution, variational Bayesian, heavy-tailed noise**

I. Introduction

The standard Gaussian approximate (GA) fixed-interval smoothers introduced in [1]–[3] are sensitive to heavy-tailed measurement noises induced by measurement outliers from unreliable sensors [4]. To solve the state estimation problem with heavy-tailed measurement noises, many robust state estimators have therefore been derived [4]–[10]. However, these robust estimators may show poor performance for heavy-tailed process noise [11].

To solve the filtering problem of linear systems with heavy-tailed process and measurement noises, Roth et al. proposed a robust Student's t filter by approximating the posterior probability density function (PDF) as Student's t [11]. However, this filter requires the growth of the degree of freedom (dof) parameters to be prevented and thereby maintain the assumption that the estimated state and process/measurement noise are jointly Student's t with a common dof parameter in the filter recursion [12]. An adaptive smoother based on a variational Bayesian (VB) approach for a linear state space model with Gaussian noises and unknown noise covariances was proposed in [13], [14], but it is sensitive to heavy-tailed process and measurement noises, as will be confirmed in Section IV. An approach to estimate the unknown parameters of a Student's t distribution for an autoregressive model was proposed in [15], however, this approach is not suitable for the state space model in this work.

This work was supported by the National Natural Science Foundation of China under Grant No. 61371173 and the Fundamental Research Funds for the Central Universities of Harbin Engineering University No. HEUCFQ20150407 and the Engineering and Physical Sciences Research Council (EPSRC) of the UK grant no. EP/K014307/1.

Y. L. Huang, Y. G. Zhang and N. Li are with the Department of Automation, Harbin Engineering University, Harbin 150001, China (e-mail: heuedu@163.com; zhangyg@hrbeu.edu.cn; ningli@hrbeu.edu.cn).

J. Chambers is with the School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK (e-mail: Jonathon.Chambers@newcastle.ac.uk).

In this letter, a robust GA fixed-interval smoother for a nonlinear state space model with heavy-tailed process and measurement noises is proposed, where the process and measurement noises are modelled as stationary Student's t distributions and the state and noise parameters are inferred approximately by using a VB approach. Simulation results show the proposed smoother outperforms existing smoothers for heavy-tailed process and measurement noises.

II. Problem formulation

Consider the following discrete-time nonlinear system

$$\begin{cases}\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1} \\ \mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k\end{cases} \tag{1}$$

where k is the discrete time index, $\mathbf{f}_{k-1}(\cdot)$ and $\mathbf{h}_k(\cdot)$ are known process and measurement functions, $\mathbf{x}_{0:T} \triangleq \{\mathbf{x}_k \in \mathbb{R}^n | 0 \leq k \leq T\}$ is the set of state vectors, and $\mathbf{z}_{1:T} \triangleq \{\mathbf{z}_k \in \mathbb{R}^m | 1 \leq k \leq T\}$ is the set of measurement vectors. The sets $\{\mathbf{w}_k \in \mathbb{R}^n | 0 \leq k \leq T-1\}$ and $\{\mathbf{v}_k \in \mathbb{R}^m | 1 \leq k \leq T\}$ contain respectively heavy-tailed process and measurement noise vectors, and they are modelled as stationary Student's t distributions as follows

$$\begin{cases}p(\mathbf{w}_k) &= \mathrm{St}(\mathbf{w}_k; \mathbf{0}, \mathbf{Q}, \omega) \\ &= \int_0^{+\infty} \mathrm{N}(\mathbf{w}_k; \mathbf{0}, \mathbf{Q}/\xi_k)\mathrm{G}(\xi_k; \frac{\omega}{2}, \frac{\omega}{2})d\xi_k \\ p(\mathbf{v}_k) &= \mathrm{St}(\mathbf{v}_k; \mathbf{0}, \mathbf{R}, \nu) \\ &= \int_0^{+\infty} \mathrm{N}(\mathbf{v}_k; \mathbf{0}, \mathbf{R}/\lambda_k)\mathrm{G}(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2})d\lambda_k\end{cases} \tag{2}$$

where $\mathrm{St}(\mathbf{w}_k; \mathbf{0}, \mathbf{Q}, \omega)$ and $\mathrm{St}(\mathbf{v}_k; \mathbf{0}, \mathbf{R}, \nu)$ denote the Student's t PDFs of $\mathbf{w}_k$ and $\mathbf{v}_k$ with mean vector $\mathbf{0}$, scale matrices $\mathbf{Q}$ and $\mathbf{R}$, and dof parameters ω and ν respectively, and $\mathrm{N}(\cdot; \mu, \mathbf{\Sigma})$ denotes the Gaussian PDF with mean vector μ and covariance matrix $\mathbf{\Sigma}$, and $\mathrm{G}(\cdot; \alpha, \beta)$ denotes the Gamma PDF with shape parameter α and rate parameter β, and ξ_k and λ_k are auxiliary random variables. The initial state vector $\mathbf{x}_0$, $\mathbf{w}_k$ and $\mathbf{v}_k$ are assumed to be mutually independent, and the initial joint PDF $p(\mathbf{x}_0, \mathbf{Q}, \omega, \mathbf{R}, \nu)$ is given as follows,

$$p(\mathbf{x}_0, \mathbf{Q}, \omega, \mathbf{R}, \nu) = \mathrm{N}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0})\mathrm{IW}(\mathbf{Q}; t_0, \mathbf{T}_0) \times \mathrm{G}(\omega; c_0, d_0)\mathrm{IW}(\mathbf{R}; u_0, \mathbf{U}_0)\mathrm{G}(\nu; a_0, b_0) \tag{3}$$

where $\mathrm{IW}(\cdot; l_0, \mathbf{L}_0)$ denotes the inverse Wishart PDF with dof parameter l_0 and inverse scale matrix $\mathbf{L}_0$, and $\hat{\mathbf{x}}_{0|0}$ and $\mathbf{P}_{0|0}$ denote respectively the initial state estimation and corresponding estimation error covariance matrix, and t_0, $\mathbf{T}_0$, c_0, d_0, u_0, $\mathbf{U}_0$, a_0 and b_0 denote respectively the prior distribution parameters of $\mathbf{Q}$, $\omega, \mathbf{R}$ and ν.

III. ROBUST GA FIXED-INTERVAL SMOOTHER

To estimate the state trajectory $\mathbf{x}_{0:T}$ of a system formulated as in (1)-(2), we need to compute the joint posterior PDF $p(\mathbf{x}_{0:T}, \mathbf{Q}, \xi_{0:T-1}, \omega, \mathbf{R}, \lambda_{1:T}, \nu|\mathbf{z}_{1:T})$, where $\xi_{0:T-1} \triangleq \{\xi_k \in \mathbb{R}|0 \leq k \leq T-1\}$ and $\lambda_{1:T} \triangleq \{\lambda_k \in \mathbb{R}|1 \leq k \leq T\}$. For a general nonlinear system, there is not an analytical solution for this posterior PDF. Thus, to obtain an approximate solution, the VB approach [16] is used to look for a free form factored approximate PDF for $p(\mathbf{x}_{0:T}, \mathbf{Q}, \xi_{0:T-1}, \omega, \mathbf{R}, \lambda_{1:T}, \nu|\mathbf{z}_{1:T})$, i.e.

$$p(\mathbf{x}_{0:T}, \mathbf{Q}, \xi_{0:T-1}, \omega, \mathbf{R}, \lambda_{1:T}, \nu|\mathbf{z}_{1:T}) \approx q(\mathbf{x}_{0:T})q(\mathbf{Q}) \times q(\xi_{0:T-1})q(\omega)q(\mathbf{R})q(\lambda_{1:T})q(\nu) \tag{4}$$

where $q(\cdot)$ is the approximate posterior PDF. According to the VB approach, these approximate posterior PDFs can be obtained by minimizing the Kullback-Leibler divergence between the approximate posterior PDF $q(\mathbf{x}_{0:T})q(\mathbf{Q})q(\xi_{0:T-1})q(\omega)q(\mathbf{R})q(\lambda_{1:T})q(\nu)$ and the true posterior PDF $p(\mathbf{x}_{0:T}, \mathbf{Q}, \xi_{0:T-1}, \omega, \mathbf{R}, \lambda_{1:T}, \nu|\mathbf{z}_{1:T})$ [17], [18], and the optimal solution satisfies the following equations

$$\log q(\phi) = \mathrm{E}_{\boldsymbol{\Theta}^{(\phi)}}[\log p(\boldsymbol{\Theta}, \mathbf{z}_{1:T})] + c_\phi \tag{5}$$

$$\boldsymbol{\Theta} \triangleq \{\mathbf{x}_{0:T}, \mathbf{Q}, \xi_{0:T-1}, \omega, \mathbf{R}, \lambda_{1:T}, \nu\} \tag{6}$$

where ϕ is an arbitrary element of $\boldsymbol{\Theta}$, and $\boldsymbol{\Theta}^{(\phi)}$ is the set of all elements in $\boldsymbol{\Theta}$ except for ϕ, and $\mathrm{E}[\cdot]$ denotes the expectation operation, and c_ϕ denotes the constant with respect to variable ϕ. Since the variational parameters of $q(\mathbf{x}_{0:T})$, $q(\mathbf{Q})$, $q(\xi_{0:T-1})$, $q(\omega)$, $q(\mathbf{R})$, $q(\lambda_{1:T})$ and $q(\nu)$ are coupled, we need to utilize fixed-point iterations to solve equation (5), where only one factor in (4) is updated while keeping other factors fixed [17].

A. Computations of approximate posterior PDFs

Using the conditional independence properties of the model (1)-(3), the joint PDF $p(\boldsymbol{\Theta}, \mathbf{z}_{1:T})$ can be factored as

$$p(\boldsymbol{\Theta}, \mathbf{z}_{1:T}) = \mathrm{N}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0})\mathrm{IW}(\mathbf{Q}; t_0, \mathbf{T}_0)\mathrm{G}(\omega; c_0, d_0)$$
$$\mathrm{IW}(\mathbf{R}; u_0, \mathbf{U}_0)\mathrm{G}(\nu; a_0, b_0)\prod_{k=1}^{T}[\mathrm{N}(\mathbf{x}_k; \mathbf{f}_{k-1}(\mathbf{x}_{k-1}), \mathbf{Q}/\xi_{k-1})$$
$$\times \mathrm{N}(\mathbf{z}_k; \mathbf{h}_k(\mathbf{x}_k), \mathbf{R}/\lambda_k)\mathrm{G}(\xi_{k-1}; \frac{\omega}{2}, \frac{\omega}{2})\mathrm{G}(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2})] \tag{7}$$

Let $\phi = \mathbf{x}_{0:T}$ and using (7) in (5), we can obtain

$$\log q^{(i+1)}(\mathbf{x}_{0:T}) = \log \mathrm{N}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0}) - 0.5\sum_{k=1}^{T}$$
$$\{[\mathbf{x}_k - \mathbf{f}_{k-1}(\mathbf{x}_{k-1})]^{\mathrm{T}}\mathrm{E}^{(i)}[\mathbf{Q}^{-1}]\mathrm{E}^{(i)}[\xi_{k-1}][\mathbf{x}_k - \mathbf{f}_{k-1}(\mathbf{x}_{k-1})]$$
$$+[\mathbf{z}_k - \mathbf{h}_k(\mathbf{x}_k)]^{\mathrm{T}}\mathrm{E}^{(i)}[\mathbf{R}^{-1}]\mathrm{E}^{(i)}[\lambda_k][\mathbf{z}_k - \mathbf{h}_k(\mathbf{x}_k)]\} + c_{\mathbf{x}} \tag{8}$$

where $(\cdot)^{\mathrm{T}}$ denotes the transpose operation, and $q^{(i+1)}(\cdot)$ is the approximation of PDF $q(\cdot)$ at the $i+1$th iteration, and $\mathrm{E}^{(i)}[\rho]$ is the expectation of variable ρ at the ith iteration. Define the modified noise covariance matrices $\tilde{\mathbf{Q}}_{k-1}^{(i)}$ and $\tilde{\mathbf{R}}_k^{(i)}$ as follows

$$\tilde{\mathbf{Q}}_{k-1}^{(i)} = \frac{\{\mathrm{E}^{(i)}[\mathbf{Q}^{-1}]\}^{-1}}{\mathrm{E}^{(i)}[\xi_{k-1}]} \quad \tilde{\mathbf{R}}_k^{(i)} = \frac{\{\mathrm{E}^{(i)}[\mathbf{R}^{-1}]\}^{-1}}{\mathrm{E}^{(i)}[\lambda_k]} \tag{9}$$

Algorithm 1: Standard GA fixed-interval smoother with modified transition and likelihood PDFs [2]

Inputs: $\mathbf{z}_{1:T}$, $\hat{\mathbf{x}}_{0|0}$, $\mathbf{P}_{0|0}$, $\tilde{\mathbf{Q}}_{k-1}^{(i)}$, $\tilde{\mathbf{R}}_k^{(i)}$

Initialization: $\hat{\mathbf{x}}_{0|0}^{(i+1)} \leftarrow \hat{\mathbf{x}}_{0|0}$, $\mathbf{P}_{0|0}^{(i+1)} \leftarrow \mathbf{P}_{0|0}$

Forward pass:

for $k = 1:T$

$\hat{\mathbf{x}}_{k|k-1}^{(i+1)} = \int \mathbf{f}_{k-1}(\mathbf{x}_{k-1})\mathrm{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}^{(i+1)}, \mathbf{P}_{k-1|k-1}^{(i+1)})d\mathbf{x}_{k-1}$

$\mathbf{P}_{k|k-1}^{(i+1)} = \int \mathbf{f}_{k-1}(\mathbf{x}_{k-1})\mathbf{f}_{k-1}^{\mathrm{T}}(\mathbf{x}_{k-1})\mathrm{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}^{(i+1)}, \mathbf{P}_{k-1|k-1}^{(i+1)}) d\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k|k-1}^{(i+1)}(\hat{\mathbf{x}}_{k|k-1}^{(i+1)})^{\mathrm{T}} + \tilde{\mathbf{Q}}_{k-1}^{(i)}$

$\mathbf{P}_{k-1,k|k-1}^{(i+1)} = \int \mathbf{x}_{k-1}\mathbf{f}_{k-1}^{\mathrm{T}}(\mathbf{x}_{k-1})\mathrm{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}^{(i+1)}, \mathbf{P}_{k-1|k-1}^{(i+1)}) d\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1|k-1}^{(i+1)}(\hat{\mathbf{x}}_{k|k-1}^{(i+1)})^{\mathrm{T}}$

$\hat{\mathbf{z}}_{k|k-1}^{(i+1)} = \int \mathbf{h}_k(\mathbf{x}_k)\mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}^{(i+1)}, \mathbf{P}_{k|k-1}^{(i+1)})d\mathbf{x}_k$

$\mathbf{P}_{zz,k|k-1}^{(i+1)} = \int \mathbf{h}_k(\mathbf{x}_k)\mathbf{h}_k^{\mathrm{T}}(\mathbf{x}_k)\mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}^{(i+1)}, \mathbf{P}_{k|k-1}^{(i+1)})d\mathbf{x}_k - \hat{\mathbf{z}}_{k|k-1}^{(i+1)}(\hat{\mathbf{z}}_{k|k-1}^{(i+1)})^{\mathrm{T}} + \tilde{\mathbf{R}}_k^{(i)}$

$\mathbf{P}_{xz,k|k-1}^{(i+1)} = \int \mathbf{x}_k\mathbf{h}_k^{\mathrm{T}}(\mathbf{x}_k)\mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}^{(i+1)}, \mathbf{P}_{k|k-1}^{(i+1)})d\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^{(i+1)}(\hat{\mathbf{z}}_{k|k-1}^{(i+1)})^{\mathrm{T}}$

$\hat{\mathbf{x}}_{k|k}^{(i+1)} = \hat{\mathbf{x}}_{k|k-1}^{(i+1)} + \mathbf{P}_{xz,k|k-1}^{(i+1)}[\mathbf{P}_{zz,k|k-1}^{(i+1)}]^{-1}[\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}^{(i+1)}]$

$\mathbf{P}_{k|k}^{(i+1)} = \mathbf{P}_{k|k-1}^{(i+1)} - \mathbf{P}_{xz,k|k-1}^{(i+1)}(\mathbf{P}_{zz,k|k-1}^{(i+1)})^{-1}(\mathbf{P}_{xz,k|k-1}^{(i+1)})^{\mathrm{T}}$

end for

Backward pass:

for $k = T:1$

$\mathbf{G}_{k-1}^{(i+1)} = \mathbf{P}_{k-1,k|k-1}^{(i+1)}[\mathbf{P}_{k|k-1}^{(i+1)}]^{-1}$

$\hat{\mathbf{x}}_{k-1|T}^{(i+1)} = \hat{\mathbf{x}}_{k-1|k-1}^{(i+1)} + \mathbf{G}_{k-1}^{(i+1)}[\hat{\mathbf{x}}_{k|T}^{(i+1)} - \hat{\mathbf{x}}_{k|k-1}^{(i+1)}]$

$\mathbf{P}_{k-1|T}^{(i+1)} = \mathbf{P}_{k-1|k-1}^{(i+1)} + \mathbf{G}_{k-1}^{(i+1)}[\mathbf{P}_{k|T}^{(i+1)} - \mathbf{P}_{k|k-1}^{(i+1)}](\mathbf{G}_{k-1}^{(i+1)})^{\mathrm{T}}$

end for

Outputs: $\{\hat{\mathbf{x}}_{k|T}^{(i+1)}, \mathbf{P}_{k|T}^{(i+1)}|0 \leq k \leq T\}$

Exploiting (8)-(9), $q^{(i+1)}(\mathbf{x}_{0:T})$ can be computed as

$$q^{(i+1)}(\mathbf{x}_{0:T}) \propto \mathrm{N}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0}) \times \prod_{k=1}^{T}[\mathrm{N}(\mathbf{x}_k; \mathbf{f}_{k-1}(\mathbf{x}_{k-1}), \tilde{\mathbf{Q}}_{k-1}^{(i)})\mathrm{N}(\mathbf{z}_k; \mathbf{h}_k(\mathbf{x}_k), \tilde{\mathbf{R}}_k^{(i)})] \tag{10}$$

It can be seen from (10) that $q^{(i+1)}(\mathbf{x}_{0:T})$ has the same form as the posterior PDF of the state in a standard nonlinear system with modified transition PDF $\mathrm{N}(\mathbf{x}_k; \mathbf{f}_{k-1}(\mathbf{x}_{k-1}), \tilde{\mathbf{Q}}_{k-1}^{(i)})$ and likelihood PDF $\mathrm{N}(\mathbf{z}_k; \mathbf{h}_k(\mathbf{x}_k), \tilde{\mathbf{R}}_k^{(i)})$. Thus, $q^{(i+1)}(\mathbf{x}_{0:T})$ can be approximated as a Gaussian PDF by using the standard GA smoother [2]. The details of the standard GA fixed-interval smoother with modified transition and likelihood PDFs are summarized in Algorithm 1 [2].

Let $\phi = \xi_{0:T-1}$ and using (7) in (5), we have

$$\log q^{(i+1)}(\xi_{0:T-1}) = \sum_{k=1}^{T}\{(\frac{n + \mathrm{E}^{(i)}[\omega]}{2} - 1)\log \xi_{k-1} - 0.5[\mathrm{E}^{(i)}[\omega] + \mathrm{tr}(\mathbf{D}_k^{(i+1)}\mathrm{E}^{(i)}[\mathbf{Q}^{-1}])]\xi_{k-1}\} + c_\xi \tag{11}$$

where $\mathrm{tr}(\cdot)$ denotes the trace operation and $\mathbf{D}_k^{(i+1)}$ is given by

$$\mathbf{D}_k^{(i+1)} = \mathrm{E}^{(i+1)}\{[\mathbf{x}_k - \mathbf{f}_{k-1}(\mathbf{x}_{k-1})][\mathbf{x}_k - \mathbf{f}_{k-1}(\mathbf{x}_{k-1})]^{\mathrm{T}}\} \tag{12}$$

Employing (11), $q^{(i+1)}(\xi_{k-1})$ can be updated as

$$q^{(i+1)}(\xi_{k-1}) = \mathrm{G}(\xi_{k-1}; \eta_{k-1}^{(i+1)}, \theta_{k-1}^{(i+1)}) \tag{13}$$

where $\eta_{k-1}^{(i+1)}$ and $\theta_{k-1}^{(i+1)}$ are given by

$$\begin{cases} \eta_{k-1}^{(i+1)} = 0.5(n + \mathrm{E}^{(i)}[\omega]) \\ \theta_{k-1}^{(i+1)} = 0.5\{\mathrm{E}^{(i)}[\omega] + \mathrm{tr}(\mathbf{D}_k^{(i+1)}\mathrm{E}^{(i)}[\mathbf{Q}^{-1}])\} \end{cases} \tag{14}$$

Let $\phi = \mathbf{Q}$ and using (7) in (5), $\log q^{(i+1)}(\mathbf{Q})$ obeys

$$\log q^{(i+1)}(\mathbf{Q}) = -0.5(t_0 + T + n + 1)\log|\mathbf{Q}| - 0.5\mathrm{tr}[(\mathbf{T}_0 + \sum_{k=1}^{T} \mathrm{E}^{(i+1)}[\xi_{k-1}]\mathbf{D}_k^{(i+1)})\mathbf{Q}^{-1}] + c_{\mathbf{Q}} \tag{15}$$

Using (15), $q^{(i+1)}(\mathbf{Q})$ can be updated as

$$q^{(i+1)}(\mathbf{Q}) = \mathrm{IW}(\mathbf{Q}; \hat{t}^{(i+1)}, \hat{\mathbf{T}}^{(i+1)}) \tag{16}$$

where $\hat{t}^{(i+1)}$ and $\hat{\mathbf{T}}^{(i+1)}$ are given by

$$\hat{t}^{(i+1)} = t_0 + T \qquad \hat{\mathbf{T}}^{(i+1)} = \mathbf{T}_0 + \sum_{k=1}^{T} \mathrm{E}^{(i+1)}[\xi_{k-1}]\mathbf{D}_k^{(i+1)} \tag{17}$$

Let $\phi = \omega$ and using (7) in (5), $\log q^{(i+1)}(\omega)$ is updated as

$$\log q^{(i+1)}(\omega) = (c_0 - 1)\log\omega - d_0\omega + \sum_{k=1}^{T}\{0.5\omega\log(0.5\omega) - \log\Gamma(0.5\omega) + (0.5\omega - 1)\mathrm{E}^{(i+1)}[\log\xi_{k-1}] - 0.5\omega\mathrm{E}^{(i+1)}[\xi_{k-1}]\} + c_\omega \tag{18}$$

where $\Gamma(\cdot)$ is the Gamma function. Using Stirling's approximation: $\log\Gamma(0.5\omega) \approx (0.5\omega - 0.5)\log(0.5\omega) - 0.5\omega$ in (18) [7], [15], $\log q^{(i+1)}(\omega)$ obeys

$$\log q^{(i+1)}(\omega) = (c_0 + 0.5T - 1)\log\omega - \{d_0 - 0.5T - 0.5\sum_{k=1}^{T}(\mathrm{E}^{(i+1)}[\log\xi_{k-1}] - \mathrm{E}^{(i+1)}[\xi_{k-1}])\}\omega \tag{19}$$

According to (19), $q^{(i+1)}(\omega)$ can be updated as

$$q^{(i+1)}(\omega) = \mathrm{G}(\omega; \hat{c}^{(i+1)}, \hat{d}^{(i+1)}) \tag{20}$$

where $\hat{c}^{(i+1)}$ and $\hat{d}^{(i+1)}$ are given by

$$\begin{cases} \hat{c}^{(i+1)} = c_0 + 0.5T \qquad \hat{d}^{(i+1)} = d_0 - 0.5T - 0.5\sum\limits_{k=1}^{T}\{ \\ \mathrm{E}^{(i+1)}[\log\xi_{k-1}] - \mathrm{E}^{(i+1)}[\xi_{k-1}]\} \end{cases} \tag{21}$$

Similar to the computation of $q^{(i+1)}(\xi_{k-1})$, let $\phi = \lambda_{1:T}$ and using (7) in (5), $q^{(i+1)}(\lambda_k)$ can be updated as

$$q^{(i+1)}(\lambda_k) = \mathrm{G}(\lambda_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}) \tag{22}$$

where $\alpha_k^{(i+1)}$ and $\beta_k^{(i+1)}$ are given by

$$\begin{cases} \alpha_k^{(i+1)} = 0.5(m + \mathrm{E}^{(i)}[\nu]) \\ \beta_k^{(i+1)} = 0.5\{\mathrm{E}^{(i)}[\nu] + \mathrm{tr}(\mathbf{E}_k^{(i+1)}\mathrm{E}^{(i)}[\mathbf{R}^{-1}])\} \end{cases} \tag{23}$$

where $\mathbf{E}_k^{(i+1)}$ is given by

$$\mathbf{E}_k^{(i+1)} = \mathrm{E}^{(i+1)}\{[\mathbf{z}_k - \mathbf{h}_k(\mathbf{x}_k)][\mathbf{z}_k - \mathbf{h}_k(\mathbf{x}_k)]^{\mathrm{T}}\} \tag{24}$$

Similar to the computation of $q^{(i+1)}(\mathbf{Q})$, let $\phi = \mathbf{R}$ and using (7) in (5), $q^{(i+1)}(\mathbf{R})$ can be updated as

$$q^{(i+1)}(\mathbf{R}) = \mathrm{IW}(\mathbf{R}; \hat{u}^{(i+1)}, \hat{\mathbf{U}}^{(i+1)}) \tag{25}$$

where $\hat{u}^{(i+1)}$ and $\hat{\mathbf{U}}^{(i+1)}$ are given by

$$\hat{u}^{(i+1)} = u_0 + T \qquad \hat{\mathbf{U}}^{(i+1)} = \mathbf{U}_0 + \sum_{k=1}^{T} \mathrm{E}^{(i+1)}[\lambda_k]\mathbf{E}_k^{(i+1)} \tag{26}$$

Likewise, for the computation of $q^{(i+1)}(\nu)$, let $\phi = \nu$ and using (7) in (5), $q^{(i+1)}(\nu)$ can be updated as

$$q^{(i+1)}(\nu) = \mathrm{G}(\nu; \hat{a}^{(i+1)}, \hat{b}^{(i+1)}) \tag{27}$$

where $\hat{a}^{(i+1)}$ and $\hat{b}^{(i+1)}$ are given by

$$\begin{cases} \hat{a}^{(i+1)} = a_0 + 0.5T \\ \hat{b}^{(i+1)} = b_0 - 0.5T - 0.5\sum\limits_{k=1}^{T}\{\mathrm{E}^{(i+1)}[\log\lambda_k] - \mathrm{E}^{(i+1)}[\lambda_k]\} \end{cases} \tag{28}$$

B. Computation of expectations

Using (13), (16), (20), (22), (25) and (27), we can compute the required expectations as follows.

$$\begin{cases} \mathrm{E}^{(i+1)}[\mathbf{Q}^{-1}] = (\hat{t}^{(i+1)} - n - 1)(\hat{\mathbf{T}}^{(i+1)})^{-1} \\ \mathrm{E}^{(i+1)}[\xi_{k-1}] = \eta_{k-1}^{(i+1)}/\theta_{k-1}^{(i+1)} \quad \mathrm{E}^{(i+1)}[\omega] = \hat{c}^{(i+1)}/\hat{d}^{(i+1)} \\ \mathrm{E}^{(i+1)}[\log\xi_{k-1}] = \psi(\eta_{k-1}^{(i+1)}) - \log\theta_{k-1}^{(i+1)} \\ \mathrm{E}^{(i+1)}[\mathbf{R}^{-1}] = (\hat{u}^{(i+1)} - m - 1)(\hat{\mathbf{U}}^{(i+1)})^{-1} \\ \mathrm{E}^{(i+1)}[\lambda_k] = \alpha_k^{(i+1)}/\beta_k^{(i+1)} \quad \mathrm{E}^{(i+1)}[\nu] = \hat{a}^{(i+1)}/\hat{b}^{(i+1)} \\ \mathrm{E}^{(i+1)}[\log\lambda_k] = \psi(\alpha_k^{(i+1)}) - \log\beta_k^{(i+1)} \end{cases} \tag{29}$$

$$\begin{cases} \mathbf{D}_k^{(i+1)} = \int\int[\mathbf{x}_k - \mathbf{f}_{k-1}(\mathbf{x}_{k-1})][\mathbf{x}_k - \mathbf{f}_{k-1}(\mathbf{x}_{k-1})]^{\mathrm{T}}\mathrm{N}(\\ \begin{bmatrix}\mathbf{x}_{k-1} \\ \mathbf{x}_k\end{bmatrix}; \begin{bmatrix}\hat{\mathbf{x}}_{k-1|T}^{(i+1)} \\ \hat{\mathbf{x}}_{k|T}^{(i+1)}\end{bmatrix}, \begin{bmatrix}\mathbf{P}_{k-1|T}^{(i+1)} & \mathbf{P}_{k-1,k|T}^{(i+1)} \\ (\mathbf{P}_{k-1,k|T}^{(i+1)})^{\mathrm{T}} & \mathbf{P}_{k|T}^{(i+1)}\end{bmatrix})d\mathbf{x}_{k-1}d\mathbf{x}_k \\ \mathbf{E}_k^{(i+1)} = \int[\mathbf{z}_k - \mathbf{h}_k(\mathbf{x}_k)][\mathbf{z}_k - \mathbf{h}_k(\mathbf{x}_k)]^{\mathrm{T}}\mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|T}^{(i+1)}, \\ \mathbf{P}_{k|T}^{(i+1)})d\mathbf{x}_k \end{cases} \tag{30}$$

where $\psi(\cdot)$ denotes the digamma function [10] and $\mathbf{P}_{k-1,k|T}^{(i+1)}$ is given by [19]

$$\mathbf{P}_{k-1,k|T}^{(i+1)} = \mathbf{G}_{k-1}^{(i+1)}\mathbf{P}_{k|T}^{(i+1)} \tag{31}$$

where $\mathbf{G}_{k-1}^{(i+1)}$ denotes the smoothing gain at the $i+1$th iteration and it is given in the fifth line from the bottom of Algorithm 1. The Gaussian weighted integrals formulated in (30) can be approximated using a sigma-point scheme, such as the third-degree spherical radial cubature rule [3]. The implementation pseudocode for the proposed robust GA fixed-interval smoother is shown in Algorithm 2, where $\mathbf{1}_{T\times 1}$ denotes the T dimensional column vector of ones.

IV. Simulation

In this section, the proposed smoother is applied to the problem of tracking an agile target which is observed by radar in clutter. The process and measurement outliers may be induced respectively by rapid motion and unreliable radar in clutter. The state-space model can be formulated as [20]

$$\mathbf{x}_k = \begin{bmatrix}\mathbf{I}_2 & \Delta t\mathbf{I}_2 \\ \mathbf{0} & \mathbf{I}_2\end{bmatrix}\mathbf{x}_{k-1} + \mathbf{w}_{k-1} \tag{32}$$

$$\mathbf{z}_k = \begin{bmatrix}\sqrt{x_k^2 + y_k^2} \\ \mathrm{atan2}(y_k, x_k)\end{bmatrix} + \mathbf{v}_k \tag{33}$$

where $\mathbf{x}_k = [x_k\ y_k\ \dot{x}_k\ \dot{y}_k]$, and x_k, y_k, $\dot{x}_k$ and $\dot{y}_k$ denote the cartesian coordinates and corresponding velocities. The parameter $\Delta t = 0.5$ is the sampling interval and $\mathbf{I}_2$ is the two

Algorithm 2

Inputs: $\mathbf{z}_{1:T}$, $\hat{\mathbf{x}}_{0|0}$, $\mathbf{P}_{0|0}$, t_0, $\mathbf{T}_0$, c_0, d_0, u_0, $\mathbf{U}_0$, a_0, b_0, N

1. Initialization: $\hat{t}^{(0)} \leftarrow t_0$, $\hat{\mathbf{T}}^{(0)} \leftarrow \mathbf{T}_0$, $\hat{c}^{(0)} \leftarrow c_0$, $\hat{d}^{(0)} \leftarrow d_0$ $\hat{u}^{(0)} \leftarrow u_0$, $\hat{\mathbf{U}}^{(0)} \leftarrow \mathbf{U}_0$, $\hat{a}^{(0)} \leftarrow a_0$, $\hat{b}^{(0)} \leftarrow b_0$, $\eta^{(0)}_{0:T-1} \leftarrow \mathbf{1}_{T\times 1}$, $\theta^{(0)}_{0:T-1} \leftarrow \mathbf{1}_{T\times 1}$, $\alpha^{(0)}_{1:T} \leftarrow \mathbf{1}_{T\times 1}$, $\beta^{(0)}_{1:T} \leftarrow \mathbf{1}_{T\times 1}$

2. Compute initial expectations using (29).

for $i = 0 : N-1$

3. Compute $\tilde{\mathbf{Q}}^{(i)}_{k-1}$ and $\tilde{\mathbf{R}}^{(i)}_{k}$ using (9).

4. Run standard GA fixed-interval smoother with modified noise covariance matrices $\tilde{\mathbf{Q}}^{(i)}_{k-1}$ and $\tilde{\mathbf{R}}^{(i)}_{k}$ in Algorithm 1.

5. Compute $\mathbf{D}^{(i+1)}_{k}$ and $\mathbf{E}^{(i+1)}_{k}$ using (30)–(31)

6. Compute $\eta^{(i+1)}_{k-1}$, $\theta^{(i+1)}_{k-1}$, $\alpha^{(i+1)}_{k}$, $\beta^{(i+1)}_{k}$ using (14) and (23).

7. Compute expectations $\mathrm{E}^{(i+1)}[\log \xi_{k-1}]$, $\mathrm{E}^{(i+1)}[\xi_{k-1}]$, $\mathrm{E}^{(i+1)}[\log \lambda_k]$, $\mathrm{E}^{(i+1)}[\lambda_k]$ using (29).

8. Compute $\hat{t}^{(i+1)}$, $\hat{\mathbf{T}}^{(i+1)}$, $\hat{c}^{(i+1)}$, $\hat{d}^{(i+1)}$, $\hat{u}^{(i+1)}$, $\hat{\mathbf{U}}^{(i+1)}$, $\hat{a}^{(i+1)}$, $\hat{b}^{(i+1)}$ using (17), (21), (26), (28)

9. Compute expectations $\mathrm{E}^{(i+1)}[\mathbf{Q}^{-1}]$, $\mathrm{E}^{(i+1)}[\omega]$, $\mathrm{E}^{(i+1)}[\mathbf{R}^{-1}]$, $\mathrm{E}^{(i+1)}[\nu]$ using (29).

end for

10. $\{\hat{\mathbf{x}}_{k|T} \leftarrow \hat{\mathbf{x}}^{(N)}_{k|T}, \mathbf{P}_{k|T} \leftarrow \mathbf{P}^{(N)}_{k|T} | 0 \le k \le T\}$

Outputs: $\{\hat{\mathbf{x}}_{k|T}, \mathbf{P}_{k|T} | 0 \le k \le T\}$

dimensional identity matrix and atan2 is the four-quadrant inverse tangent function. Similar to [11], outlier corrupted process and measurement noises are generated according to

$$\begin{cases} \mathbf{w}_k \sim \begin{cases} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_w) & \text{w.p. } 0.8 \\ \mathrm{N}(\mathbf{0}, 1000\boldsymbol{\Sigma}_w) & \text{w.p. } 0.2 \end{cases} \\ \mathbf{v}_k \sim \begin{cases} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_v) & \text{w.p. } 0.8 \\ \mathrm{N}(\mathbf{0}, 100\boldsymbol{\Sigma}_v) & \text{w.p. } 0.2 \end{cases} \end{cases} \tag{34}$$

where w.p. denotes "with probability" and $\boldsymbol{\Sigma}_w$ and $\boldsymbol{\Sigma}_v$ are nominal process and measurement noise covariance matrices

$$\boldsymbol{\Sigma}_w = \begin{bmatrix} \frac{\Delta t^3}{3}\mathbf{I}_2 & \frac{\Delta t^2}{2}\mathbf{I}_2 \\ \frac{\Delta t^2}{2}\mathbf{I}_2 & \Delta t \mathbf{I}_2 \end{bmatrix} \quad \boldsymbol{\Sigma}_v = \begin{bmatrix} 100\mathrm{m}^2 & 0 \\ 0 & 16\mathrm{mrad}^2 \end{bmatrix} \tag{35}$$

In this simulation, the standard cubature Kalman smoother (CKS) [2], outlier robust CKS [4], CKS with unknown noise covariances (CKSWUNC) [13], [14], the proposed robust CKS with fixed noise parameters (the proposed CKS-fixed), the proposed robust CKS with estimated $\mathbf{Q}$ and $\mathbf{R}$ and fixed ω and ν (the proposed CKS-$\mathbf{QR}$), the proposed robust CKS with estimated ω and ν and fixed $\mathbf{Q}$ and $\mathbf{R}$ (the proposed CKS-$\omega\nu$), and the proposed robust CKS with estimated $\mathbf{Q}$, $\mathbf{R}$, ω and ν (the proposed CKS-$\mathbf{QR}\omega\nu$) are tested. Note that CKSWUNC is obtained by using the Rauch-Tung-Striebel smoother in [13] combined with the third degree spherical radial cubature rule [3] based statistical linearization of the nonlinear system. The scale matrix and dof parameter of the existing outlier robust CKS are set as $\boldsymbol{\Sigma}_v$ and 5. The parameters of existing CKSWUNC are set as: $\nu_0 = 6$, $\mathbf{V}_0 = \boldsymbol{\Sigma}_w$, $\mu_0 = 4$, $\mathbf{M}_0 = \boldsymbol{\Sigma}_v$. In the proposed robust CKS, the initial parameters of estimated noise parameters are set as: $t_0 = 6$, $\mathbf{T}_0 = \boldsymbol{\Sigma}_w$, $u_0 = 4$, $\mathbf{U}_0 = \boldsymbol{\Sigma}_v$, $a_0 = c_0 = 5$, $b_0 = d_0 = 1$, and the fixed noise parameters $\mathbf{Q}$, $\mathbf{R}$, ω and ν are respectively set as $\boldsymbol{\Sigma}_w$, $\boldsymbol{\Sigma}_v$, 5, and 5. The initial true state vector $\mathbf{x}_0 = [10000, 1000, 300, -40]^T$, and the initial estimation error covariance matrix $\mathbf{P}_{0|0} = \mathrm{diag}([100\ 100\ 100\ 100])$, and the initial state estimation $\hat{\mathbf{x}}_{0|0}$ is chosen randomly from $N(\mathbf{x}_0, \mathbf{P}_{0|0})$. The number of measurements is chosen as $T = 200$, and the number of variational iteration is chosen as $N = 10$, and 1000 independent Monte Carlo runs are performed. The root-mean square errors (RMSEs) of position and velocity are chosen as performance metrics, which are defined as

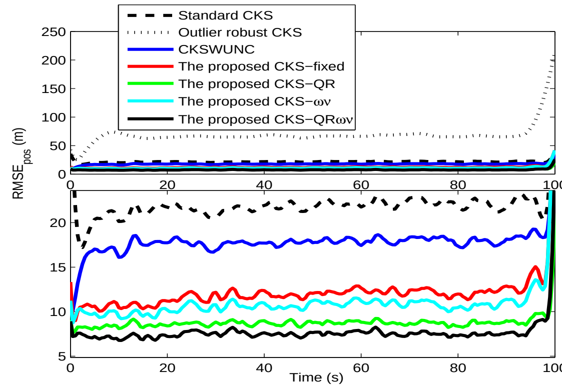

Fig. 1: RMSE of position.

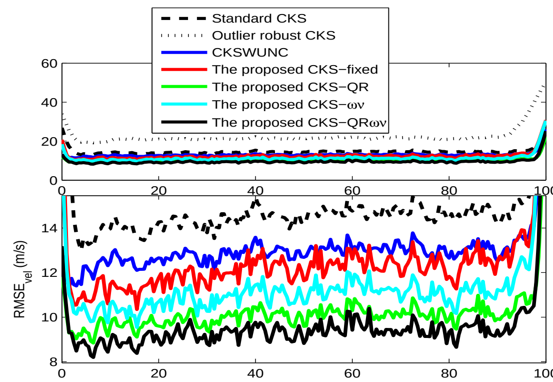

Fig. 2: RMSE of velocity.

$$\mathrm{RMSE}_{\mathrm{pos}} = \sqrt{\frac{1}{M}\sum_{s=1}^{M}[(x^s_k - \hat{x}^s_{k|k})^2 + (y^s_k - \hat{y}^s_{k|k})^2]} \tag{36}$$

where (x^s_k, y^s_k) and $(\hat{x}^s_{k|k}, \hat{y}^s_{k|k})$ are the true and estimated positions at the s-th Monte Carlo run and M denotes the number of Monte Carlo runs. Similar to the RMSE in position, we can also write formula for the RMSE in velocity.

Fig.1-Fig. 2 respectively show the RMSEs of position and velocity from the proposed CKSs and existing CKSs. It can be seen from Fig.1-Fig. 2 that RMSEs from the proposed CKSs are smaller than that from existing CKSs. We can also see from Fig.1-Fig. 2 that both the proposed CKS-$\mathbf{QR}$ and the proposed CKS-$\omega\nu$ have smaller RMSEs than the proposed CKS-fixed, and the proposed CKS-$\mathbf{QR}\omega\nu$ has the smallest RMSEs. Thus, the estimation accuracy of the proposed smoother is further improved by learning noise parameters adaptively from data.

V. Conclusion

In this letter, a robust GA fixed-interval smoother for nonlinear systems with heavy-tailed process and measurement noises was derived based on the VB approach. The simulation results of radar tracking with process and measurement outliers showed the proposed smoother has better estimation accuracy than existing GA fixed-interval smoothers.

REFERENCES

[1] S. Särkkä, "Unscented Rauch-Tung-Striebel Smoother," *IEEE Trans. Autom. Control*, vol. 53, no. 3, pp. 845−849, Apr 2008.

[2] S. Särkkä and J. Hartikainen, "On Gaussian optimal smoothing of non-linear state space models," *IEEE Trans. Autom. Control*, vol. 55, no. 8, pp. 1938–1941, Aug 2010.

[3] I. Arasaratnam and S. Haykin, "Cubature Kalman smoothers," *Automatica*, vol. 47, no. 10, pp. 2245−2250, Oct 2011.

[4] R. Piché, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution," *in Proceedings of MLSP*, Sept. 2012.

[5] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "An outlier-robust Kalman filter," *in IEEE International Conference on Robotics and Automation (ICRA)*, May 2011, pp. 1551–1558.

[6] G. Agamennoni, J.I. Nieto, and E.M. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5024–5037, Oct. 2012.

[7] H. Zhu, H. Leung, and Z. He, "A variational Bayesian approach to robust sensor fusion based on Student-t distribution," *Inf. Sci.*, vol. 221, no. 2013, pp. 201–214, Sep. 2012.

[8] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson, "Robust inference for state-space models with skewed measurement noise," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 2450–2454, Dec. 2015.

[9] G. Agamennoni and E.M. Nebot, "Robust estimation in non-linear state-space models with state-dependent noise," *IEEE Trans. Signal Process.*, vol. 62, no. 8, Apr. 2014.

[10] D. J. Xu, C. Shen, and F. Shen, "A robust particle filtering algorithm with non-Gaussian measurement noise using Student-t distribution," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 30–34, Jan 2014.

[11] M. Roth, E. Özkan, and F. Gustafsson, "A Student's t filter for heavy-tailed process and measurement noise," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 5770–5774.

[12] S. Saha, "Noise robust online inference for linear dynamic systems," [Online]. Available: http://arxiv.org/pdf/1504.05723

[13] T. Ardeshiri, E. Ozkan, U. Orguner, F. Gustafsson, "Approximate Bayesian smoothing with unknown process and measurement noise covariances," *IEEE Signal Process. Lett.*, vol.22, no.12, pp. 2450–2454, Dec. 2015.

[14] T. Ardeshiri, E. Ozkan, U. Orguner, "Approximate Bayesian smoothing with unknown process and measurement noise covariances," arXiv:1412.5307v1

[15] J. Christmas, R. Everson, "Robust autoregression: Student-t innovations using variational Bayes," *IEEE Trans. Signal Process.*, vol.59, no.1, pp. 48–57, Jan. 2011.

[16] S. Särkkä, and J. Hartikainen, "Variational Bayesian adaptation of noise covariance in nonlinear Kalman filtering," *arXiv: 1302.0681v1 [Stat.ME]*, Feb 2013.

[17] C. M. Bishop, *Pattern Recognition and Machine Learning.* Springer, 2007.

[18] D. Tzikas, A. Likas, and N. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.

[19] J. Kokkala, A. Solin, and S. Särkkä, "Expectation maximization based parameter estimation by sigma-point and particle smoothing," in *Proceedings of the 17th International Conference on Information Fusion*, Salamanca, Spain, July 2014, pp. 1–8.

[20] I. Arasaratnam, S. Haykin, and R. Elliott, "Discrete-time nonlinear filtering algorithms using Gauss-Hermite quadrature," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953−977, May 2007.